

**K-MEANS CLUSTERING DAN EXTREME GRADIENT  
BOOSTING (XGBOOST) UNTUK PREDIKSI POIN  
DI GAME FANTASY PREMIER LEAGUE**

**SKRIPSI**

Diajukan untuk Memenuhi Sebagian dari  
Syarat Memperoleh Gelar Sarjana Komputer  
Program Studi Ilmu Komputer



Oleh  
Yuliansyah Ibrahim  
NIM 1701988

PROGRAM STUDI ILMU KOMPUTER  
DEPARTEMEN PENDIDIKAN ILMU KOMPUTER  
FAKULTAS PENDIDIKAN MATEMATIKA DAN ILMU PENGETAHUAN ALAM  
UNIVERSITAS PENDIDIKAN INDONESIA  
2022

**K-MEANS CLUSTERING DAN EXTREME GRADIENT  
BOOSTING (XGBOOST) UNTUK PREDIKSI POIN  
DI GAME FANTASY PREMIER LEAGUE**

Oleh:

Yuliansyah Ibrahim

1701988

Sebuah Skripsi yang Diajukan untuk Memenuhi Salah Satu Syarat Memperoleh  
Gelar Sarjana Komputer di Fakultas Pendidikan Matematika dan Ilmu  
Pengetahuan Alam

© Yuliansyah Ibrahim 2022

Universitas Pendidikan Indonesia

Juni 2022

Hak Cipta Dilindungi Undang-Undang

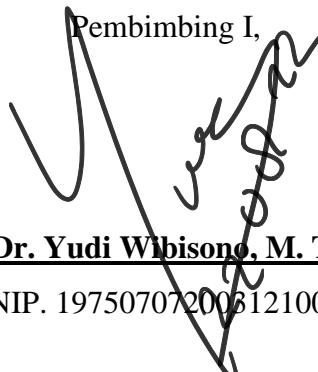
Skripsi ini tidak boleh diperbanyak seluruhnya atau sebagian, dengan dicetak  
ulang, difoto kopi atau cara lainnya tanpa izin dari penulis

YULIANSYAH IBRAHIM

1701988

**K-MEANS CLUSTERING DAN EXTREME GRADIENT  
BOOSTING (XGBOOST) UNTUK PREDIKSI POIN  
DI GAME FANTASY PREMIER LEAGUE**

DISETUJUI DAN DISAHKAN OLEH PEMBIMBING:

Pembimbing I,  
  
Dr. Yudi Wibisono, M. T.  
NIP. 197507072003121003  
Pembimbing II,



Dr. Rani Megasari, M. T.  
NIP. 198705242014042002

Mengetahui,  
Ketua Program Studi Ilmu Komputer



Dr. Rani Megasari, M. T.  
NIP. 198705242014042002

## **PERNYATAAN**

Dengan ini saya menyatakan bahwa skripsi dengan judul “K-Means Clustering dan Extreme Gradient Boosting (XGBoost) untuk Prediksi Poin di Game Fantasy Premier League” ini beserta seluruh isinya adalah benar-benar karya saya sendiri. Saya tidak melakukan penjiplakan atau pengutipan dengan cara-cara yang tidak sesuai dengan etika ilmu yang berlaku dalam masyarakat keilmuan. Atas pernyataan ini, saya siap menanggung risiko atau sanksi apabila di kemudian hari ditemukan adanya pelanggaran etika keilmuan atau ada klaim dari pihak lain terhadap keaslian karya saya ini.

Palembang, Juni 2022  
Yang membuat pernyataan,



Yuliansyah Ibrahim

NIM 1701988

**K-MEANS CLUSTERING DAN EXTREME GRADIENT  
BOOSTING (XGBOOST) UNTUK PREDIKSI POIN  
DI GAME FANTASY PREMIER LEAGUE**

Oleh

Yuliansyah Ibrahim — yuliansyahibr@upi.edu

1701988

**ABSTRAK**

Popularitas *Fantasy Sport* di kalangan penggemar olahraga professional semakin meningkat setiap tahunnya. Fantasy Premier League (FPL) merupakan game fantasy sport resmi Liga Premier Inggris yang dimainkan oleh jutaan orang di seluruh dunia. Partisipan FPL berperan sebagai manajer sebuah tim yang bertujuan untuk memperoleh poin sebanyak-banyaknya. Kemampuan analisis dan pengambilan keputusan sangat diandalkan dalam menentukan pemain dalam tim yang dimiliki. *Platform* yang menyediakan model untuk memprediksi poin pemain FPL sudah banyak tersedia di pasar. Dengan sifatnya yang dinamis dan sulit ditebak, bidang prediksi poin pada Fantasy Premier League terbuka lebar untuk peningkatan. Pada penelitian ini digunakan metode penggabungan model *unsupervised learning* K-Means dengan model *supervised learning* Extreme Gradient Boosting (XGBoost). Data yang digunakan adalah empat musim data Fantasy Premier League mulai dari musim 2017 hingga musim 2020. Data yang terkumpul diolah dan diagregasi, lalu digabungkan menjadi satu dataset utuh. Berdasarkan eksperimen yang dilakukan, ditemukan bahwa penggabungan K-Means Clustering dengan XGBoost tidak meningkatkan kinerja model. Model tunggal XGBoost menghasilkan nilai RMSE 2.118 dan MAE 1.176. Sedangkan model *ensemble* XGBoost dan K-Means pada nilai parameter k terbaik menghasilkan nilai RMSE 2.431 dan MAE 1.945.

Kata kunci — fantasy premier league, fpl, fantasy sport, clustering regression, cluster ensemble, xgboost, k-means clustering, prediksi poin

## KATA PENGANTAR

Puji syukur saya panjatkan kepada Allah Subhanahu wa Ta'ala atas rahmat dan karunia-Nya serta shalawat dan salam kepada Nabi Muhammad ﷺ, sehingga penyusunan proposal skripsi ini dapat diselesaikan dengan baik dan tepat waktu dari yang telah ditentukan. “K-Means Clustering dan Extreme Gradient Boosting (XGBoost) untuk Prediksi Poin di Game Fantasy Premier League” ini dapat dilaksanakan dengan baik karena bantuan, bimbingan, dan kerja sama dari berbagai pihak. Oleh karena itu, ucapan terima kasih penulis sampaikan kepada semua pihak yang telah membantu penyusunan proposal skripsi ini.

Saya menyadari bahwa dalam penyusunan proposal ini masih banyak kekurangan, oleh karena itu saran dan kritik yang membangun sangat saya harapkan. Semoga laporan ini memberikan manfaat dan kebaikan pada mahasiswa yang akan menyusun skripsi di tahun yang akan datang.

Palembang, Juni 2022



Penyusun

Yuliansyah Ibrahim

## UCAPAN TERIMAKASIH

Alhamdulillah, puji dan syukur penulis panjatkan kepada ke hadirat Allah Ta'ala yang telah melimpahkan rahmat serta hidayah-Nya, sehingga dalam penulisan dan penelitian ini diberikan kelancaran dan dapat selesai. Tentu pada saat proses penelitian dan penulisan ini penulis mendapatkan dorongan, semangat, bimbingan, dan juga bantuan dari berbagai pihak. Oleh karena itu penulis mengucapkan rasa terima kasih yang sedalam-dalamnya serta penghargaan yang tinggi kepada:

1. Orang tua penulis yang sudah membantu dan memberikan dorongan serta doa kepada penulis.
2. Bapak Dr. Yudi Wibisono M. T. selaku pembimbing I, dan Ibu Dr. Rani Megasari, M. T. selaku pembimbing II dan juga Ketua Program Studi Ilmu Komputer yang telah meluangkan waktu untuk membimbing penelitian dan penulisan skripsi ini.
3. Bapak dan Ibu Dosen Program Studi Pendidikan Ilmu Komputer dan Ilmu Komputer yang telah memberikan ilmu yang bermanfaat kepada penulis selama masa perkuliahan.
4. Teman-teman yang telah membantu proses penelitian dan mendukung penulis dalam melakukan penelitian.
5. Seluruh pihak lainnya yang tidak dapat penulis sebutkan satu persatu yang telah membantu dalam proses penelitian.

Semoga semua amal baik dari perbuatan dan doa yang telah diberikan oleh orang-orang yang berperan kepada penulis mendapatkan balasan yang berlipat ganda dari Allah. Aamiin.

Palembang, Juni 2022



Yuliansyah Ibrahim

## DAFTAR ISI

ABSTRAK .....	i
KATA PENGANTAR .....	ii
UCAPAN TERIMAKASIH.....	iii
DAFTAR ISI.....	iv
DAFTAR GAMBAR .....	vii
DAFTAR TABEL.....	viii
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang .....	1
1.2 Rumusan Masalah .....	3
1.3 Tujuan Penelitian .....	3
1.4 Manfaat Penelitian .....	3
1.5 Batasan Masalah.....	4
1.6 Sistematika Penulisan .....	4
BAB II KAJIAN PUSTAKA.....	6
2.1 <i>Machine Learning</i> .....	6
2.1.1 <i>Supervised Learning</i> .....	6
2.1.2 <i>Unsupervised Learning</i> .....	7
2.2 Proses <i>Machine Learning</i> .....	7
2.3 <i>Fantasy Sport</i> .....	9
2.4 Fantasy Premier League.....	9
2.4.1 Aturan Permainan.....	10
2.4.2 Sistem poin .....	12
2.4.3 Kompetisi dan hadiah.....	14
2.5 Extreme Gradient Boosting.....	16
2.5.1 Dasar Teori XGBoost.....	16
2.5.2 Regression Tree .....	16
2.5.3 Teknik <i>Tree Boosting</i> .....	19
2.5.4 Algoritma XGboost .....	21
2.6 <i>Library XGBoost Python</i> .....	28
2.7 <i>K-Means Clustering</i> .....	28

2.8	<i>Ensemble Learning</i> dengan XGBoost dan <i>K-means Clustering</i> .....	31
2.9	Evaluasi Model.....	33
2.9.1	Elbow Method .....	33
2.9.2	Mean Squared Error (MSE) dan Root Mean Squared Error (RMSE)....	34
2.9.3	Mean Absolute Error (MAE) .....	35
	BAB III METODOLOGI PENELITIAN .....	36
3.1	Desain Penelitian.....	36
3.1.1	Studi Literatur.....	36
3.1.2	Pengumpulan Data .....	36
3.1.3	Analisis .....	37
3.1.4	Perancangan.....	38
3.1.5	Pembangunan .....	38
3.1.6	Eksperimen .....	42
3.1.7	Pengujian .....	43
3.1.8	Analisis Hasil .....	44
3.1.9	Dokumentasi.....	44
3.2	Alat dan Bahan Penelitian.....	44
	BAB IV TEMUAN DAN PEMBAHASAN.....	45
4.1	Pengumpulan Data .....	45
4.2	Praproses Data.....	49
4.2.1	Penanganan File Data yang Tidak Lengkap.....	50
4.2.2	Pemilihan Fitur .....	52
4.2.3	Modifikasi dan Ekstraksi Fitur .....	56
4.2.4	Penyatuan data.....	61
4.3	Eksperimen.....	62
4.3.1	Model Ensemble K-Means dan XGBoost .....	63
4.3.2	Pembangunan Model Ensemble K-Means dengan Model Lainnya .....	68
4.3.3	Eksperimen Menggunakan Dataset Tanpa Fitur Aggregasi .....	72
4.4	Analisis Hasil Eksperimen .....	73
	BAB V KESIMPULAN DAN SARAN.....	75
5.1	Kesimpulan .....	75
5.2	Saran.....	75



## DAFTAR GAMBAR

Gambar 2.1 Pengisian tim setelah pendaftaran .....	11
Gambar 2.2 Contoh Perolehan Poin di FPL.....	13
Gambar 2.3 Contoh pohon klasifikasi.....	17
Gambar 2.4 Contoh pohon regresi .....	18
Gambar 2.5 Algoritma Gradient Boosting .....	20
Gambar 2.6 Pohon hasil pemecahan di simpul akar .....	24
Gambar 2.7 Pohon hasil pemecahan di simpul kanan.....	25
Gambar 2.8 Bentuk akhir pohon yang dihasilkan.....	26
Gambar 2.9 Gambaran model pada iterasi pertama .....	26
Gambar 2.10 Hasil akhir model .....	27
Gambar 2.11 Visualisasi contoh data untuk K-Means Clustering. ....	31
Gambar 2.12 Diagram rancangan alur metode (Trivedi, Pardos, & Heffernan, 2011) .....	32
Gambar 3.1 Desain penelitian .....	36
Gambar 3.2 Diagram rancangan implementasi.....	37
Gambar 4.1 Repository github data Fantasy Premier League.....	45
Gambar 4.2 Struktur direktori data Fantasy Premier League. ....	47
Gambar 4.3 <i>Keyword</i> Pencarian Data Lama di archive.org.....	51
Gambar 4.4 Hasil pencarian arsip API Fantasy Premier League .....	51
Gambar 4.5 Data API Fantasy Premier League .....	52
Gambar 4.6 Konversi data API ke dalam format CSV .....	52
Gambar 4.7 Jumlah total pemain per posisi .....	54
Gambar 4.8 Rata-rata poin per posisi.....	54
Gambar 4.9 Daftar 67 fitur dataset yang digunakan .....	62
Gambar 4.10 Diagram <i>flowchart</i> proses perhitungan rata-rata prediksi akhir model <i>ensemble</i> .....	65
Gambar 4.11 Plot hasil perhitungan metode Elbow.....	66

## DAFTAR TABEL

Tabel 2.1 Sebaran harga pemain berdasarkan posisi di musim 2020 .....	12
Tabel 2.2 Aturan Perolehan Poin di FPL .....	14
Tabel 2.3 Daftar hadiah yang dapat diraih manajer FPL .....	15
Tabel 2.4 Contoh data karyawan yang digunakan .....	22
Tabel 2.5 Nilai residual prediksi awal.....	22
Tabel 2.6 Hasil pencarian <i>best split</i> pada simpul akar .....	23
Tabel 2.7 Hasil pencarian <i>best split</i> pada simpul kanan .....	24
Tabel 2.8 Hasil pencarian <i>best split</i> pada simpul kanan .....	25
Tabel 2.9 Hasil perhitungan nilai residual baru .....	26
Tabel 2.10 Hasil prediksi menggunakan model final.....	28
Tabel 2.11 Contoh data untuk K-means clustering.....	30
Tabel 2.12 Titik centroid dari <i>cluster</i> 1 dan <i>cluster</i> 2 .....	30
Tabel 2.13 Perhitungan jarak ke <i>cluster</i> 1 dan 2 pada iterasi pertama.....	31
Tabel 3.1 Fitur data <i>team.csv</i> .....	39
Tabel 3.2 Fitur data <i>fixtures.csv</i> .....	39
Tabel 3.3 Fitur data <i>merged_gw.csv</i> .....	40
Tabel 4.1 Tabel contoh sampel data <i>teams.csv</i> .....	46
Tabel 4.2 Tabel contoh sampel data <i>fixtures.csv</i> .....	47
Tabel 4.3 Tabel contoh sampel data <i>merged_gw.csv</i> .....	48
Tabel 4.4 Tabel status ketersedian data.....	50
Tabel 4.5 Jumlah data Fantasy Premier League.....	53
Tabel 4.6 Daftar fitur yang digunakan dari setiap file .....	55
Tabel 4.7 Contoh sampel data hasil pergeseran indeks.....	57
Tabel 4.8 Sampel data hasil perhitungan kumulatif.....	58
Tabel 4.9 Sampel data perhitungan kumulatif musim sebelumnya .....	59
Tabel 4.10 Sampel data perhitungan <i>sum</i> musim sebelumnya .....	60
Tabel 4.11 Sampel data perhitungan jumlah pertandingan.....	61
Tabel 4.12 Parameter model XGBoost. ....	63
Tabel 4.13 Pseudo-code pembangunan model <i>ensemble</i> .....	64
Tabel 4.14 Pseudo-code prediksi model <i>ensemble</i> .....	66

Tabel 4.15 Sampel hasil prediksi model .....	67
Tabel 4.16 Kinerja eksperimen model <i>ensemble</i> K-Means dan XGBoost.....	68
Tabel 4.17 Parameter yang digunakan pada model Random Forest, Decision Tree, dan Gaussian Naïve Bayes.....	69
Tabel 4.18 Kinerja eksperimen model <i>ensemble</i> Random Forest.....	69
Tabel 4.19 Kinerja eksperimen model <i>ensemble</i> Decision Tree.....	70
Tabel 4.20 Kinerja eksperimen model <i>ensemble</i> Gaussian Naïve Bayes .....	71
Tabel 4.21 Kinerja model eksperimen <i>ensemble</i> XGBoost pada dataset tanpa fitur agregasi. ....	73
Tabel 4.22 Rekap hasil pengukuran kinerja model dari semua eksperimen .....	74

## DAFTAR PUSTAKA

- Albon, C. (2018). *Machine learning with python cookbook: Practical solutions from preprocessing to deep learning.* " O'Reilly Media, Inc.".
- Anand, V. (2021). Fantasy-Premier-League [Sumber data]. <https://github.com/vaastav/Fantasy-Premier-League>
- Andrews, Z. (2019). *Comparing Predictive Models for English Premier League Games.* (Disertasi). Appalachian State University.
- Arnas, R. (2014). *English Premier League As The United Kingdom's Soft Power Instrument.* (Disertasi). President University.
- Bonello, N., Beel, J., Lawless, S., & Debattista, J. (2019). Multi-stream Data Analytics for Enhanced Performance Prediction in Fantasy Football. arXiv preprint arXiv:1912.07441.
- Burton, R., Hall, K., & Paul, R. (2013). The historical development and marketing of fantasy sports leagues. *The Journal of Sport*, 2(2), 185-215.
- Carlson, C. (2013). The reality of fantasy sports: A metaphysical and ethical analysis. *Journal of the Philosophy of Sport*, 40(2), 187-204.
- Chen, T., & Guestrin, C. (2016, August). Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining* (pp. 785-794).
- Coors, S. (2018). *Automatic gradient boosting.* (Tesis). Ludwig-Maximilians-Universitat Munchen.
- Dubber J. (2015). *Playing the game: The soft power of sport.* [Online]. Diakses dari: <https://www.britishcouncil.org/research-policy-insight/insight-articles/playing-game-soft-power-sport>
- Dwyer, B. & Kim, Y. (2011). For Love or Money: Developing and Validating a Motivational Scale for Fantasy Football Participation. *Journal of Sport Management*. 25. 70-83. 10.1123/jsm.25.1.70.
- Freund, Y., & Schapire, R. E. (1996, July). "Experiments with a new boosting algorithm". Dalam *Proceedings of the Thirteenth International Conference on International Conference on Machine Learning* (Vol. 96, hlm. 148-156), San Francisco: Morgan Kaufmann Publishers Inc.

- Goldberg, D. E., & Holland, J. H. (1988). Genetic algorithms and machine learning.
- Gupta, A. (2019). Time Series Modeling for Dream Team in Fantasy Premier League. arXiv preprint arXiv:1909.12938.
- Kristiansen, B. K., Gupta, A., & Eilertsen, W. (2018). *Developing a Forecast-Based Optimization Model for Fantasy Premier League* (Master's thesis, NTNU).
- Lutz, R. (2015). Fantasy football prediction. arXiv preprint arXiv:1505.06918.
- Marshland, S. (2015). *Machine Learning an algorithmic perspective*. Chapman and Hall/CRC.
- Mitchell, T. (1997). *Machine learning*. McGraw-Hill Professional.
- Mohammed, M., Khan, M. B., & Bashier, E., B., M. (2017). *Machine Learning : Algorithms and Applications*. CRC Press.
- Mohri, M., Rostamizadeh, A., & Talwalkar, A. (2018). *Foundations of machine learning*. MIT press.
- O'Brien, J. D., Gleeson, J. P., & O'Sullivan, D. J. (2020). Identification of skill in an online game: The case of Fantasy Premier League. arXiv preprint arXiv:2009.01206.
- Qian, G., Wu, Y., Ferrari, D., Qiao, P., & Hollande, F. (2016). “Semisupervised clustering by iterative partition and regression with neuroscience applications”. *Computational Intelligence and Neuroscience* 2016 (hlm. 814-826).
- Raschka, S., & Mirjalili, V. (2019). *Python Machine Learning : Machine Learning and Deep Learning with Python, Scikit-learn, and TensorFlow 2, 3rd Edition*. Packt Publishing.
- Shipman, F. M. (2001, April). “Blending the real and virtual: Activity and spectatorship in fantasy sports”. Dalam *Proceedings from DAC'01: The Conference on Digital Arts and Culture* (Vol. 8, hlm. 2007).
- Trivedi, S., Pardos, Z. A., & Heffernan, N. T. (2011). “Clustering students to generate an ensemble to improve standard test score predictions”. Dalam *International conference on artificial intelligence in education* (hlm. 377-384). Springer, Berlin, Heidelberg.
- Trivedi, S., Pardos, Z. A., & Heffernan, N. T. (2015). “The utility of clustering in prediction tasks”.

- Touzani, S., Granderson, J., & Fernandes, S. (2018). “Gradient boosting machine for modeling the energy consumption of commercial buildings”. Dalam *Energy and Buildings* (vol. 158, hlm. 1533-1543).
- Sammut, C., & Webb, G. I. (2017). *Encyclopedia of Machine Learning and Data Mining*. Springer.
- Spinda, J. S., & Havard, C. (2016). “I Wouldn’t Pick Them To Save My Season”. Dalam *Fantasy Sports and the Changing Sports Media Industry: Media, Players, and Society* (hlm. 63).